Dataset Construction

LLM

[Motion Caption]: The camera **remains static while yawing right**, **stays static**, and then **moves forward**.

LLM

[Motion Tags]: (translation + rotation)
Frames 0-47: static + yaw right
Frames 48-67: static + static
Frames 68-119: move forward + static

[Directorial Caption]: (**Movements** + **Scene** + **Intent**)
The camera **starts static, yawing right** to **capture the extensive area** and reveal **more equipment** as it pivots. The shot continues to **hold steady**, allowing for **clear observation of the surroundings**, including **dirt inclines and the still water reflecting nearby trees**. The camera then **smoothly moves forward** to the water's edge, keeping focus on the **calm water surface**.

Pre-processing and Trajectory extraction

time

Motion Tagging

Trajectory Generation & Applications

Multi-modal Input | **Trajectory Gen** | **Camera-Controlled Video**

Text

[Motion Caption] The camera **stays static while rolling right**, then **moves forward continuing to roll right**.

Text + RGBD

[Directorial Caption] The camera moves forward **along the valley**, showcasing the rolling hills and yellow fields.

+ A sunflower swaying in the wind.